РОЗДІЛ 2

Що технологія не покращить

> *Тренувальні дані для машинного навчання без еталонної істини — це шлях до провалу. Те, що модель може забезпечити, починається лише там, де істина задокументована.*

Є надія, яка з'являється майже в кожному AI-проекті. Її рідко висловлюють, але вона структурує очікування всіх учасників. Надія звучить так: якщо модель буде достатньо великою, вона виправить погані вхідні дані.

Не виправить. Йдеться передусім про кероване навчання, де модель тренується на парах «вхід → правильна відповідь» (типовий випадок індустріального AI: контроль якості, передбачувальне обслуговування, кредитний скоринг). За поточних доказів останніх десяти років жодна промислова модель цього класу сама по собі структурно цього не вирішила, і немає інженерних підстав очікувати, що наступне покоління це зробить. Те, що дає машинне навчання, — це апроксимація функції, закладеної в тренувальних даних. Якщо тренувальні дані суперечливі, модель апроксимує суперечності. Якщо неповні, апроксимує прогалини. Якщо просто хибні, апроксимує хибну реальність — точно, швидко і в індустріальному масштабі.

McKinsey у дослідженні *State of AI 2025* зафіксував співвідношення, що задає масштаб проблеми: 78% організацій використовують AI щонайменше в одній бізнес-функції, але лише 1% описують себе як *зрілі в AI* — тобто такі, у яких AI реально вбудована в операційне ядро і дає вимірювану віддачу. Розрив між *використовуємо* і *зріло використовуємо* — не питання технології. Це питання людського шару, еталонної істини і організаційної архітектури навколо моделі.

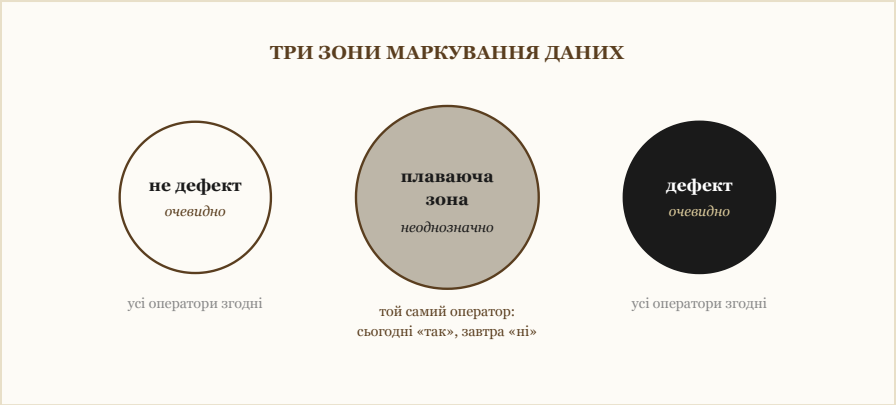
Кейт Кроуфорд у праці *Atlas of AI* показує, що питання даних — це питання класифікації: те, що в наборі даних вважається *дефектом* або *нормою*, — це кодифікований організаційний вибір, що носить вигляд нейтрального технічного факту. Якість даних залишається постійною функцією з власним ритмом перевірки, виходячи далеко за межі одноразового аудиту перед запуском, регламентом і призначеним відповідальним. Без цього розмова про еталонну істину швидко зводиться до технічних деталей налаштування моделі (підбору параметрів, які впливають на швидкість і характер навчання) — тобто до питань, які не вирішують проблеми, що лежить рівнем нижче.

Що таке еталонна істина насправді

Еталонна істина (буквально «істина на місці») — це задокументована правильна відповідь, відносно якої можна виміряти модель. На виробництві це означає: дефект ідентифіковано, категоризовано і позначено так, щоб друга особа з тим самим інструментом дійшла до тієї самої позначки.

Сама собою ця вимога звучить просто. Але саме вона стає головним вузьким місцем будь-якого індустріального проекту: відтворюваність позначення.

Звідки взагалі береться еталонна істина? З процедури маркування даних: хтось дивиться на конкретний образ — фотографію поверхні — і фіксує: дефект чи ні, який саме. На очевидних випадках усе просто — явна тріщина, явно чиста поверхня, тут оператори-маркувальники дають однакові відповіді. Проблема починається на неочевидних: дефект на межі допуску, неоднозначна структура, прикордонний випадок.



Три зони. Білі і чорні випадки — стабільне маркування. Сіра зона — джерело всіх подальших проблем: модель навчається на «плаваючій правді» і повторює це нестабільне сприйняття на конвеєрі.

У більшості індустріальних проектів дефекти на межі цієї сірої зони не позначені відтворювано. Те, що сьогодні класифіковано як тріщину, завтра — дефект лакофарбового покриття. Те, що одна зміна позначає як брак, наступна приймає. Цей різнобій у маркуванні — не виняток, а норма. Він виникає системно: дефекти на межі допуску виглядають візуально подібно, і жодна людина не тримає однакове сприйняття цілодобово.

Наслідок — ефект каскаду: на різних об'єктах, у різних клієнтів та на іншому обладнанні система працює по-різному; рекламації ростуть; команда підтримки тоне в нескінченних розборах окремих випадків: «у нас спрацювало, у них — ні»; собівартість підтримки росте швидше, ніж економія від автоматизації. У певний момент проект фактично починає коштувати дорожче, ніж той ручний контроль, який він замінив.

Чому збільшення моделі тут не допомагає

На таких даних можна тренувати модель — будь-якого розміру і архітектури. Вона пройде процес навчання і в кінці видасть формальний показник точності. Цей показник описує лише те, наскільки добре модель повторила суперечливі позначки людей-маркувальників на тестовій вибірці. Він не описує, наскільки добре вона насправді розпізнає дефекти у виробництві.

У крайовому випадку модель залишається такою ж ненадійною, як і людина, що її навчала. Крайовий випадок — це дефект, який оператор-маркувальник сьогодні фіксує, а завтра пропускає. Саме там модель мала б допомогти. Саме там вона не справляється.

> *У більшості індустріальних контекстів більше тренувальних даних мало що покращує, якщо дані суперечливі. Більші моделі мало що покращують, якщо сам сигнал нечіткий. Складніші архітектури мало що покращують, якщо те, що треба апроксимувати, не визначене стабільно.*

Випадок із виробництва: як виглядає сіра зона в реальному проекті

Один з таких випадків — розробка оптичних приладів для високоточних промислових вимірювань. Перші моделі, навчені на наявному наборі даних з людськими позначками, давали порядку 95 % точності на тестовій вибірці. На реальному виробничому потоці саме крайові випадки з сірої зони генерували помилки, які зупиняли калібрування і поверталися як рекламації. Після того, як постановку задачі переформу-

лювали (про що далі), та сама архітектура моделі на тих самих реальних знімках вийшла на 98,5 % — і, що важливіше, давала відтворюваний результат між різними запусками, а не «плавала» від навчання до навчання.

Рішення було не в більшій моделі. Воно було в переформулюванні самої постановки задачі. Замість того, щоб змушувати модель відтворювати людську непослідовність, було задокументовано фізичні властивості дефекту: форму, глибину, характер відбиття світла. Ці властивості можна задавати в симуляції. Із симуляції народжуються синтетичні зображення дефектів з відомими наперед параметрами — тобто еталонна істина визначена не постфактум людиною, а в момент створення кожного прикладу.

У пілотному проекті навчили двадцять моделей суто на таких синтетичних даних — кожну з різним випадковим стартом. Усі двадцять давали узгоджені передбачення на тих самих реальних виробничих знімках: різниця між більшістю моделей — у межах статистичного шуму. Жодна модель, навчена на людських мітках із сірої зони, такої стабільності не давала. Двадцять моделей збігалися між собою саме тому, що реальність, до якої вони підганялися, була стабільно задокументована — а не «плавала» від зміни до зміни.

Коли синтетичні дані допомагають

Описаний випадок — окремий приклад ширшої схеми. Синтетичні дані дозволяють те, що не дає жодне нарощування людського маркування: вони виносять еталонну істину з-під людської непослідовності. Кожен приклад має відому істину за побудовою. На таких даних можна тренувати моделі довільної архітектури і чесно порівнювати їх між собою; можна моделювати ситуації, які в реальному виробництві трапляються

раз на рік, і навчати модель на них так само щільно, як на буденних; можна цілеспрямовано підсилювати ті випадки, на яких модель помиляється, не чекаючи природного потрапляння таких прикладів у конвеєр.

Де синтетичні дані не допомагають

Синтетичні дані — не універсальне рішення. Вони працюють там, де фізичне явище зрозуміле і відтворюване в симуляції. Не працюють у доменах, де сама істина ще не відкрита, наприклад у складних соціальних даних або в явищах, механіка яких не моделюється.

Що вони все ж роблять: переносять вузьке місце з добування даних на моделювання явища. Хто може описати фізичну поведінку дефекту, той може створити скільки завгодно тренувальних прикладів. Хто не може — той ніколи й не мав справжньої еталонної істини. Машинне навчання проект би не врятувало.

Висновок

Не кожна проблема даних — це проблема даних. Деякі — проблеми визначення. Хто замовляє AI-проект, не перевіривши, чи взагалі існує еталонна істина, той замовляє дороге повторення того, що виробництво вже має: ту саму непослідовність, у вищому темпі, з рахунком за розробку, інтеграцію і додаткове обладнання для обчислень (GPU) зверху.

Наступний розділ присвячений другій ролі антропології: хто контролює, що саме робить модель у роботі.

· · ·

Три питання для власного контексту

1. **Чи існує відтворювана еталонна істина?** Чи дійшли б два незалежних оператори-маркувальники даних з тими самими даними до тих самих позначок? Чи перевірено це вибіркою на 100 випадках, чи лише припускається?

2. **Якщо ні: чи моделюється сам дефект (або інше явище, яке має розпізнавати модель) у фізичному сенсі?** Чи можна задокументувати його фізичні, правилами описувані або симулятивні властивості так, щоб генерувати синтетичні дані з відомою істиною?

3. **Якщо й це неможливо: чи машинне навчання — правильний важіль?** Чи це проблема визначення, яку треба вирішити організаційно перед будь-яким моделюванням?